NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

ON DISTRIBUTION-FREE TOLERANCE LIMITS IN RANDOM SAMPLING

By Herbert Robbins

Post Graduate School, Annapolis, Md.

Let $X_1, \cdots, X_n$ be independent random variables each with the continuous and differentiable cumulative distribution function $\sigma(x) = Pr(X_i < x)$. A continuous function $f(x_1, \cdots, x_n)$ with the property that the random variable $Y = \sigma(f(X_1, \cdots, X_n))$ has a probability distribution which is independent of $\sigma(x)$ will be called a *distribution-free upper tolerance limit*[1] (d. f. u. t. l.). We shall prove

THEOREM 1. *A necessary and sufficient condition that the continuous function $f(x_1, \cdots, x_n)$ be a d. f. u. t. l. is that the function*

$$\bar{f}(x_1, \cdots, x_n) = \prod_{i=1}^{n} \{f(x_1, \cdots, x_n) - x_i\}$$

be identically zero.

PROOF. Since f is continuous, we can prove the necessity of the condition by deriving a contradiction from the assumption that f is a d. f. u. t. l. for which there exist distinct numbers $a_1, \cdots, a_n$ such that $f(a_1, \cdots, a_n) = A \neq a_i$, $(i = 1, \cdots, n)$.

Since the numbers $a_1, \cdots, a_n, A$ are distinct, there will exist a positive number ϵ such that the $(n + 1)$ intervals

$$I: \quad A - \epsilon \leq x \leq A + \epsilon$$

$$I_i: \quad a_i - \epsilon \leq x \leq a_i + \epsilon \qquad (i = 1, \cdots, n),$$

have no points in common. Moreover, since f is continuous, there will correspond to ϵ a positive number $\epsilon_1 < \epsilon$ such that

$$A - \epsilon \leq f(x_1, \cdots, x_n) \leq A + \epsilon,$$

provided that simultaneously

$$|x_i - a_i| < \epsilon_1 \qquad (i = 1, \cdots, n).$$

Now let p be any number between $\frac{1}{3}$ and $\frac{2}{3}$. Corresponding to p we define the function $\sigma_p(x)$ as follows. In the interval I we set $\sigma_p(x) = p$. In every interval

$$J_i: \quad a_i - \epsilon_1 \leq x \leq a_i + \epsilon_1 \qquad (i = 1, \cdots, n)$$

[1] Cf. S. S. Wilks, *Mathematical Statistics*, Princeton University Press (1943), pp. 93–94.

we let $\sigma_p(x)$ increase an amount $\left(\frac{1}{3n}\right)$. Outside the intervals $I, J_1, \cdots, J_n$ we define $\sigma_p(x)$ in any manner so that it is continuous, differentiable, and non-decreasing for every x, and has the properties $\sigma_p(-\infty) = 0$, $\sigma_p(\infty) = 1$. It is clear that we can do this.

Let S denote the set of all points $(x_1, \cdots, x_n)$ of n dimensional space such that simultaneously

$$|x_i - a_i| \leq \epsilon_1 \qquad (i = 1, \cdots, n).$$

Then by construction, for $\sigma_p(x)$ defined above,

$$Pr((X_1, \cdots, X_n) \in S) = \left(\frac{1}{3n}\right)^n.$$

But if $(X_1, \cdots, X_n) \in S$, then by construction,

$$A - \epsilon \leq f(X_1, \cdots, X_n) \leq A + \epsilon$$

and

$$Y = \sigma_p(f(X_1, \cdots, X_n)) = p.$$

Hence for $\sigma(x) = \sigma_p(x)$ we have

$$Pr(Y = p) \geq \left(\frac{1}{3n}\right)^n.$$

But since f is a d. f. u. t. l., this inequality must hold for any $\sigma(x)$.

Now choose a set of numbers

$$\tfrac{1}{3} < p_1 < p_2 < \cdots < p_m < \tfrac{2}{3},$$

where $m = 2(3n)^n$. Then from the above,

$$Pr(Y = \text{one of the numbers } p_1, \cdots, p_m) \geq 2.$$

This is the desired contradiction.

Let $O_r(x_1, \cdots, x_n)$ be the function whose value is the rth term when the numbers $x_1, \cdots, x_n$ are arranged in non-decreasing order of magnitude. In terms of the functions O_r we can characterize the continuous functions f which satisfy the identity $\bar{f} \equiv 0$ as follows. Let $i_1, \cdots, i_n$ be a permutation of the integers $1, \cdots, n$. Denote by $E(i_1, \cdots, i_n)$ the set of all points $(x_1, \cdots, x_n)$ such that

$$x_{i_1} < x_{i_2} < \cdots < x_{i_n}.$$

The $n!$ sets E are open and disjoint. Since f is continuous and $\bar{f} \equiv 0$, in each $E(i_1, \cdots, i_n)$ we must have, for some r,

$$f(x_1, \cdots, x_n) \equiv O_r(x_1, \cdots, x_n),$$

where the integer $r = r(i_1, \cdots, i_n)$ must depend on the permutation $i_1, \cdots, i_n$ in such a way that f may be extended continuously over the whole space. (The condition for this is as follows. Two permutations $i_1, \cdots, i_n$ and $j_1, \cdots, j_n$ may be called *adjacent* if they differ only by an interchange of two adjacent integers. Then for any two adjacent permutations, either $r(i_1, \cdots, i_n) = r(j_1, \cdots, j_n)$ or the two values of r are the two interchanged integers. For example, the function

$$f(x_1, x_2, x_3) = \begin{cases} O_3(x_1, x_2, x_3) & \text{if } O_3(x_1, x_2, x_3) = x_1 \\ O_2(x_1, x_2, x_3) & \text{otherwise} \end{cases}$$

satisfies this requirement.)

We shall now prove that the necessary condition, $\bar{f} \equiv 0$, of Theorem 1 is sufficient to ensure that the continuous function f be a d. f. u. t. l. From the argument of the preceding paragraph, any continuous function f such that $\bar{f} \equiv 0$ will in each set $E(i_1, \cdots, i_n)$ have the value $O_r(x_1, \cdots, x_n)$, where r is an integer from 1 to n. Since the variables $X_1, \cdots, X_n$ are independent and have the same probability distribution, the probability that $(X_1, \cdots, X_n)$ will belong to $E(i_1, \cdots, i_n)$ is equal to $(1/n!)$ for every permutation $i_1, \cdots, i_n$. Let

$$W = f(X_1, \cdots, X_n).$$

Then if $\varphi(x) = d\sigma(x)/dx$ denotes the probability density function of each X_i, the conditional p. d. f. of $W = O_r(X_1, \cdots, X_n)$, given that $(X_1, \cdots, X_n)$ belongs to $E(i_1, \cdots, i_n)$, will be $n!\psi_r(w)$, where

$$\psi_r(w) = \frac{\varphi(w)\sigma^{r-1}(w)[1 - \sigma(w)]^{n-r}}{(r-1)!\,(n-r)!}.$$

Thus $\psi_r(w)$ will be of the form

$$\psi_r(w) = \varphi(w)F_r(\sigma(w)),$$

where $F_r(\sigma(w))$ is a polynomial in $\sigma(w)$. Hence the conditional p. d. f. of $Y = \sigma(W)$, given that $(X_1, \cdots, X_n)$ belongs to $E(i_1, \cdots, i_n)$, will be $n!\xi_r(y)$, where

$$\xi_r(y) = F_r(y),$$

and the p. d. f. of Y will be

$$\xi(y) = \Sigma F_r(y),$$

where the summation is over the $n!$ integers $r = r(i_1, \cdots, i_n)$. This is independent of $\sigma(x)$, so that f is a d. f. u. t. l. This completes the proof of Theorem 1.

A function $f(x_1, \cdots, x_n)$ is *symmetric* if its value is unchanged by any permutation of its arguments. It is clear that the only continuous and symmetric functions f which satisfy the identity $\bar{f} \equiv 0$ are the n functions $O_r(x_1, \cdots, x_n)$. Hence we can state

THEOREM 2. *The only symmetric d. f. u. t. l.'s are the n functions* $O_r(x_1, \cdots, x_n)$ $(r = 1, \cdots, n)$.